Journal of Student Research Exploration

https://shmpublisher.com/index.php/josre
p-ISSN 2964-1691 | e-ISSN 2964-8246

Comparison of KNN, Naive Bayes, and Decision Tree methods in predicting the accuracy of classification of Immunotherapy dataset

Nadhifah Nur Annisa Reska[1], Khansa Tsabita[2]
[1, 2, 3]Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received June 14, 2023
Revised July 6, 2023
Accepted July 7, 2023

Keywords:

Immunotherapy
Naive bayes
Decision tree
K-nearest neighbor

ABSTRACT

Health is crucial for humans to carry out daily activities, and cancer is the second leading cause of death worldwide. Maintaining health is essential in minimizing factors associated with cancer. Immunotherapy is a new cancer treatment technique that has s shown a bigger success rate compared with conventional techniques. However, the effectiveness of this method depends on accurate diagnosis, which requires deeper analysis and research on classification methods. This study compares the accuracy of KNN, Naive Bayes, and Decision Tree classification methods in predicting the accuracy of immunotherapy treatment. The goal is to find the most effective classification techniques that can provide more accurate predictive results in treating diseases using immunotherapy. Based on the test results of Naive Bayes, Decision Tree, and K-Nearest Neighbor, the result obtained of accuracy rates are 81.11%, 80.00%, and 74.44%. From the accuracy comparison, it is known that the Naive Bayes algorithm is the most effective algorithm with the highest accuracy value of 81.11%.

1. Introduction

Health is essential for humans in carrying out various activities. The human physical condition will experience a decrease and obstacles if Clarity. Without

[1]***Corresponding Author:***
Nadhifah Nur Annisa Reska,
Department of Computer Science,
Universitas Negeri Semarang,
Sekaran, Gunungpati, Semarang, Indonesia.
Email: nadhifa@students.unnes.ac.id
DOI: https://doi.org/10.52465/josre.v1i2.170

health, humans will be influenced by genetic, environmental, and lifestyle factors such as eating, drinking, working, resting, and regulating emotions [1]. One of the biggest health issues in the globe is cancer. According to the Global Burden of Disease, 9.56 million individuals passed away from cancer before in 2017. Cancer is the sixth leading cause of mortality worldwide [2]. The death data can be seen in Figure 1.

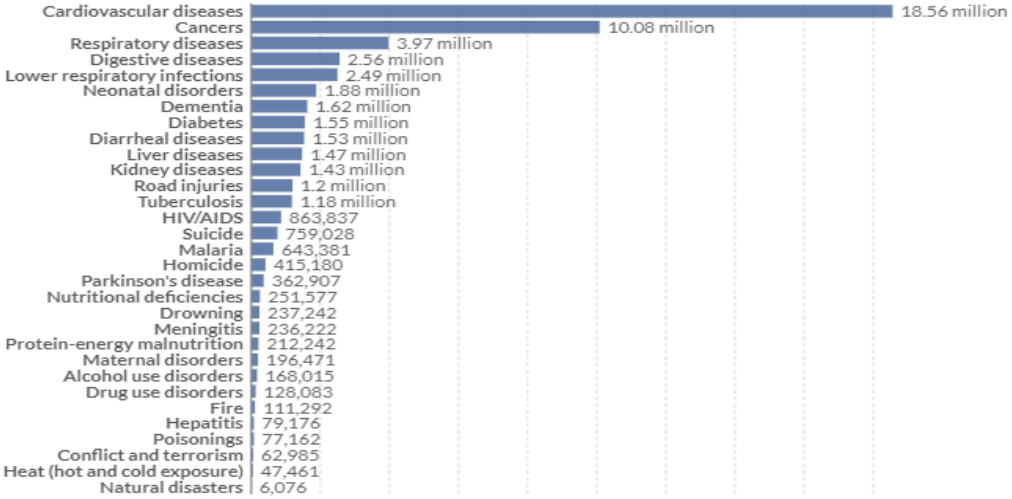


Figure 1. Number of deaths by cause (Source: IHME, Global Burden of Disease (2019))

Figure 1 is a diagram of death data in which cancer is the 2nd most common cause of death from various causes of death.

According to Riskesdas statistics, Indonesia's tumor/cancer prevalence increased from 1.4 per 1000 people in 2013 to 1.79 per 1000 people in 2018. In the meanwhile, according to statistics from the World Health Organization (WHO)'s Global Burden of Cancer Study (Globocan), there were 396,914 cancer diagnoses and 234,511 deaths in Indonesia in 2020. The high cases of cancer can certainly be our initial vigilance to minimize the factors associated with this cancer. This affects the development of research in finding new drugs, even from natural ingredients now widely studied for the treatment of this cancer [3].

Maintaining health is very important for human life. Cancer patients often feel burdened by the long and expensive treatment process [4]. Therefore technology is something that is needed by humans [5] and an effective and successful treatment technology is needed in treating cancer to reduce patient anxiety and burden.

One of the cancer treatments is Immunotherapy. An innovative treatment method called immunotherapy has been identified to treat skin cancer, and it has an opportunity to solve the issues with the Cryotherapy approach [6]. In a study [7] the results show that Immunotherapy has a higher success compared to Cryotherapy. The study stated that a significant therapeutic response was found in

patients undergoing Immunotherapy compared to Cryotherapy, where 76.7% of patients fully healed with the Immunotherapy method and 56.7% of patients fully healed with the Cryotherapy method.

The data used is immunotherapy data, taken from the UCI Machine Learning Repository immunotherapy. This dataset is made up of continuous and nominal numeric attributes. Due to the data's endless structure, continuous characteristics might result in lesser accuracy. As a result, discretization is required to transform characteristics into discrete data [8].

The effectiveness of the Immunotherapy method also depends heavily on the correct and accurate diagnosis of the disease the patient is facing. Therefore, it is necessary to carry out more in-depth analysis and research regarding the classification method used in predicting the accuracy of Immunotherapy [9]. Classification is data mining technique to make predictions about classes within a dataset [10]. Several classification methods that are often used in data processing are KNN (K-Nearest Neighbors), Naive Bayes, and Decision Tree [11]. Each approach to predicting classification has benefits and drawbacks.

The KNN (K-Nearest Neighbor) algorithm is one of the popular algorithms used in data classification[11]. The KNN method has the advantage that it is very effective when applied to large amounts of data and to training data that has a lot of noise. However, the KNN method also has weaknesses, namely the lack of optimization in setting K values and in determining the best attribute to be selected [12].

In addition to the KNN method, researchers also use the Decision Tree method. This algorithm is usually used in statistical pattern recognition [13]. The Decision Tree consists of three nodes, namely the root node which is the starting point of a decision tree, an intermediate node related to a test, and a leaf node [14].

Apart from KNN, the Decision Tree in providing predictive results also uses the Naive Bayes method which is a machine learning method for probability [15]. This method is often used for text classification with high processing speed on large data. Naive Bayes is also used for predictions because it contains simple probabilistic which are applied to Bayes' theorem with strong dependence [14].

This study aims to compare the accuracy of the predictions produced by the three classification methods in the Immunotherapy dataset. Through this research, it is hoped that the most effective classification method can be found and can provide more accurate predictive results in the treatment of diseases treated by immunotherapy.

2. Method

K-Nearest Neighbor, Naive Bayes, and Decision Tree are classification methods utilized by the authors in this study to classify the Immunotherapy Dataset. This research was conducted to make a comparison between K-Nearest Neighbor, Naive Bayes, and Decision Tree to see whether method is superior to the two methods in predicting the accuracy of the success of the immunotherapy treatment method. The stages of the research conducted can be seen in Figure 2.

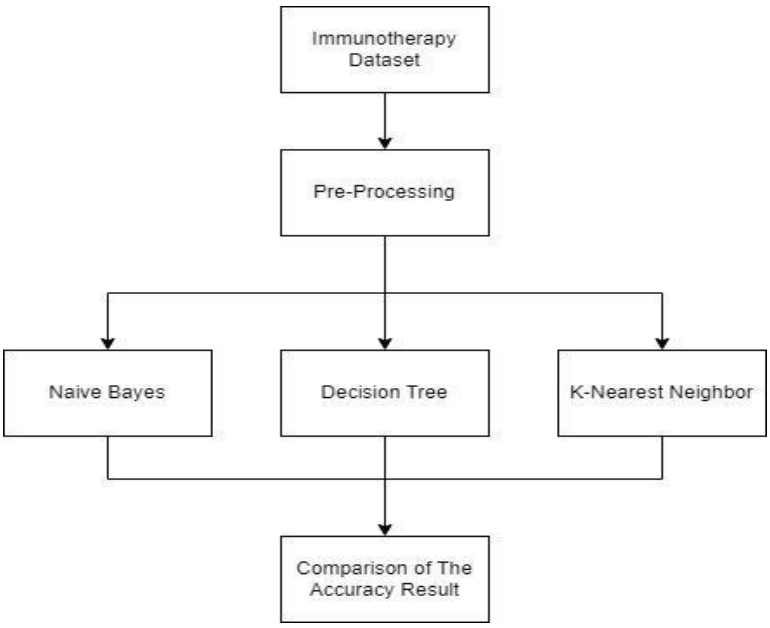


Figure 2. Research stages

In Figure 2, the initial stage in the research conducted was to prepare a dataset regarding Immunotherapy to be used in the research. After that, the second stage is pre-processing where the data will be normalized to eliminate empty data that might reduce model performance. After the pre-processing process, the data will be processed using Naive Bayes, Decision Tree, and K-Nearesr Neighbor and then the results of the algorithm research will be compared to find the most effective algorithm.

2.1.Immunotherapy Dataset

Immunotherapy is a dataset that contains information about the results of the cancer treatment of 90 patients using the Immunotherapy treatment method sourced from the UCI Machine Learning Repository. Several previous related studies used the Immunotherapy dataset, namely research conducted by Umri Erdiansyah et.al in 2021 with the research title, "Comparison of the K-Nearest Neighbor and Random Forest Methods in Predicting the Accuracy of Classification of Wart Disease Treatment". The number of data records in the dataset is 90 data records with a total of 7 attributes and 1 attribute class [16]. The attribute in datasets includes sex, age, time, number of warts, type, area, induration diameter, and result of treatment.

2.2.K-Nearest Neighbor

Instance-based learning is a subset of which K-Nearest Neighbor (KNN) is one. This algorithm may alternatively be viewed as a lazy learning technique. To find groups of k items in the training data that are identical to or like the objects in the testing data or fresh data, the K-nearest neighbor technique must be used. When classifying an item, KNN compares the learning data that is closest to the object [17]. By measuring the similarity between new cases and old cases, the K-Nearest Neighbor approach itself provides a way to locate cases. This is accomplished by comparing the weights of several existing traits. You may use the Euclidean formula to calculate the separation between two points, namely the point on the training data (x) and the point on the testing data (y) [18].

$$D_{(x_1,x_2)} = \sqrt{\sum_{i=1}^{n}(x_1 - x_2)^2} \quad (1)$$

D : Closest distance.

x_1 : Data training.

x_2 : Data testing.

n : Numerous characteristics for each situation.

i : Individual characteristics from 1 to n.

The quantity of data that is accurately anticipated is known as accuracy, and it may be determined using the formula below.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (2)$$

TP (True Positives) = The number of positive objects correctly classified.

TN (True Negatives) = The number of negative objects misclassified.

FP (False Positives) = The number of negative objects correctly classified.

FN (False Negatives) = The number of positive objects misclassified [19].

2.3.Naive Bayes

The algorithm known as Naive Bayes may be used to categorize data. A statistical technique called Bayesian classification forecasts the likelihood of belonging to a class. The word "naive" is paired with Bayes' theorem to denote the independence of each characteristic or variable. In supervised learning, naive Bayes may be effectively learnt. The benefit of this classifier is that it only needs a little amount of training data to estimate the classification parameters, such as the mean and variance of variables. Only the variation of the variable for each class has to be computed since independent variables are assumed, not the whole covariance matrix. Naive Bayes makes the assumption during the classification process that the existence or absence of a feature in a class is unrelated to the existence of other

features in the same class. When performing classification, the Bayesian approach will produce the category label with the highest probability [20].

$$P(X|Y) = \frac{P(Y|X)(x)}{P(Y)} \quad (3)$$

$P(X|Y)$ = probability X probability value based on Y conditions.

$P(Y|X)$ = the probability Y that X determines is true.

$P(X)$ = probability of X value.

$P(Y)$ = probability of Y value [21].

2.4.Decision Tree (C4.5 Algorithm)

Decision tree is a quite efficient method of classifying data [22]. The decision tree is a method of classification that makes use of a tree structure. Each node on the tree corresponds to an attribute, and each branch to the attribute's value. The tree's leaves stand in for categories or labels. To find solutions to a given problem, decision trees are employed as a method of deductive reasoning. The resulting tree's shape is not necessarily binary. The tree will be in the shape of a binary tree if all the features in the data set have just two category values. The resultant tree, however, is typically not a binary tree if there are more than two category entries or if a numeric data type is used. The leaf nodes indicate classes, whereas each interior node represents a variable.

A collection of decision tree methods known as Algorithm C4.5 are used to construct decision trees. A group of algorithms known as the C4.5 algorithm are used in data mining and machine learning to solve classification issues. The ID3 algorithm was improved upon to create the C4.5 algorithm. The C4.5 and ID3 algorithms were first created by J. Ross Quinlan, an artificial intelligence researcher, in the late 1970s. Algorithm C4.5 constructs the decision tree from top to bottom, with the top attribute acting as the leaf and the bottom attribute as the root [23].

The stages in calculating the C4.5 algorithm include preparing training data, determining the root of the tree by calculating entropy, calculating gain, and then determining the tuples to be partitioned. The formulas for the C4.5 algorithm below.

$$Entropy\ (S) = \sum_{i=1} -pi(\log 2\ pi) \quad (4)$$

S : case set.

A : feature.

n : number of partitions S.

pi : the proportion of Si to S.

$$Gain\ (S, A) = Entropy\ (S)\ - \sum_{i=1}^{n} \frac{|S_i|}{|S|} (Entropy\ (S_i)) \quad (5)$$

S : case set.

A : feature.

n : number of partitions S.

Si : number of cases on the i partition [24].

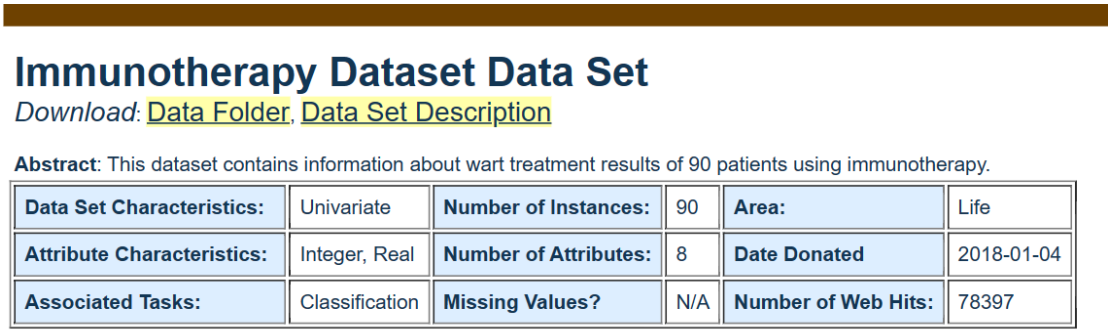
3. Results and Discussion

The Naïve Bayes and K-Nearest Neighbor tools used are the RapidMiner applications. Here the study uses the Immunotherapy data set taken from the UCI repository (University of California, Irvene) [25] with 90 data records. The attribute in datasets includes sex, age, time, number of warts, type, area, induration diameter, and result of treatment.

3.1.Testing Stage

In this study, we used the Naive Bayes, Decision Tree, and K-Nearest Neighbor methods which are divided into several stages, namely:

1. Download the Immunotherapy dataset on the UCI website (University of California, Invene). An overview of the data used in this study can be seen in Figure 3.



Immunotherapy Dataset Data Set

Download: Data Folder, Data Set Description

Abstract: This dataset contains information about wart treatment results of 90 patients using immunotherapy.

Data Set Characteristics:	Univariate	**Number of Instances:**	90	**Area:**	Life
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	8	**Date Donated**	2018-01-04
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	78397

Figure 3. Screenshot of the data source web view

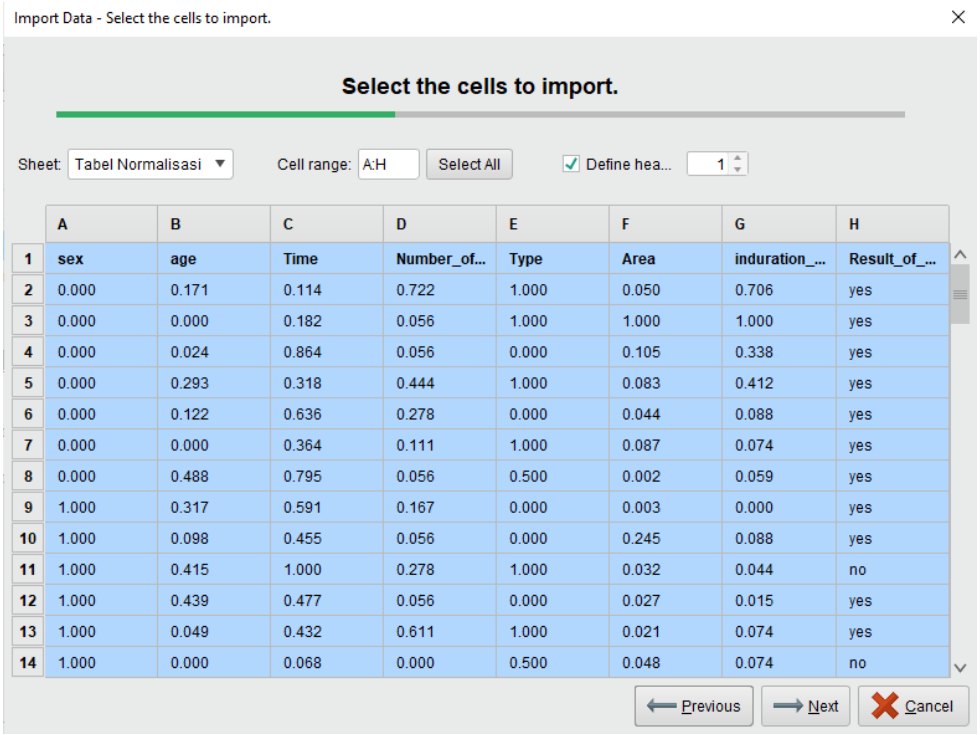
2. Change the variable from the result of the treatment attribute which starts from numeric (0 & 1) to char (yes & no). Then normalize the characteristics in the dataset to be tested by applying the min-max normalization technique. Min-Max normalization normalizes data to a minimum value equal to 0 and a maximum value equal to 1. The equation formula used for min-max normalization is in the following equation.

$$N* = \frac{N - min(n)}{max} \quad (6)$$

$N*$: Data normalization results.

$min(n)$: The minimum value of the attribute of the matter is to be determined.

3. After the data normalization process is complete, then the testing process is carried out using the rapid miner application. Figure 4 is an illustration of the data normalization process carried out.



	A	B	C	D	E	F	G	H
1	sex	age	Time	Number_of...	Type	Area	induration_...	Result_of_...
2	0.000	0.171	0.114	0.722	1.000	0.050	0.706	yes
3	0.000	0.000	0.182	0.056	1.000	1.000	1.000	yes
4	0.000	0.024	0.864	0.056	0.000	0.105	0.338	yes
5	0.000	0.293	0.318	0.444	1.000	0.083	0.412	yes
6	0.000	0.122	0.636	0.278	0.000	0.044	0.088	yes
7	0.000	0.000	0.364	0.111	1.000	0.087	0.074	yes
8	0.000	0.488	0.795	0.056	0.500	0.002	0.059	yes
9	1.000	0.317	0.591	0.167	0.000	0.003	0.000	yes
10	1.000	0.098	0.455	0.056	0.000	0.245	0.088	yes
11	1.000	0.415	1.000	0.278	1.000	0.032	0.044	no
12	1.000	0.439	0.477	0.056	0.000	0.027	0.015	yes
13	1.000	0.049	0.432	0.611	1.000	0.021	0.074	yes
14	1.000	0.000	0.068	0.000	0.500	0.048	0.074	no

Figure 4. Normalized data import

Form Figure 4, dataset must be entered first then pre-processing can be done.

4. After that enter the operator's component according to the method used. An overview of the process can be seen in Figure 5.

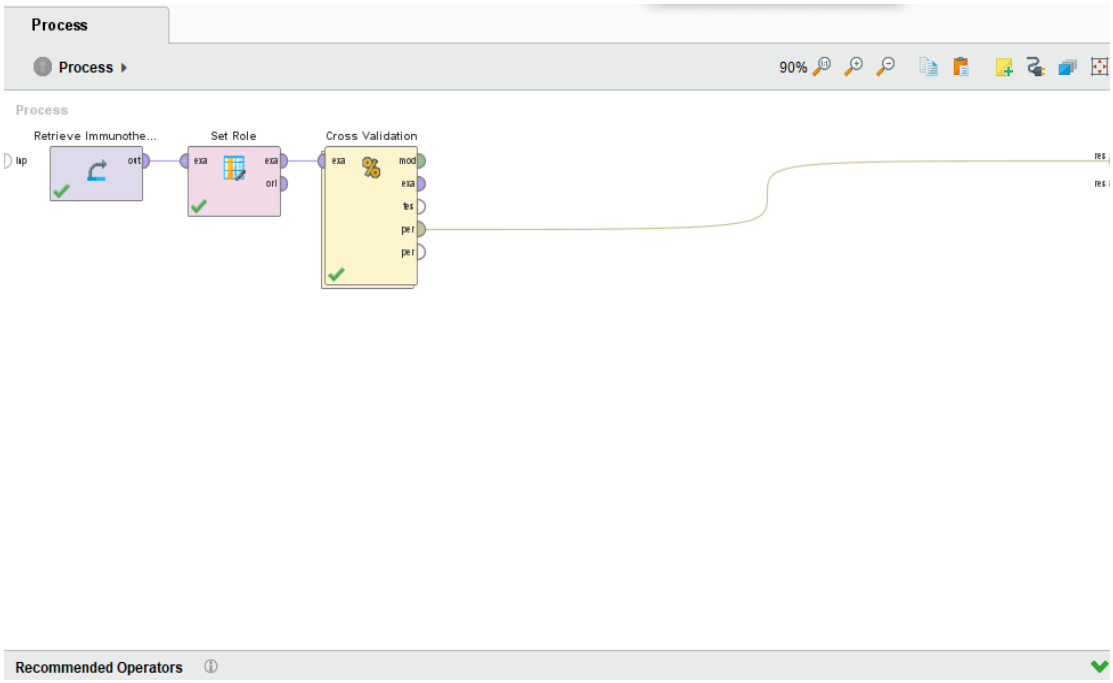


Figure 5. Sub-Process

Figure 5 is an example of the initial sub-process that was carried out on the Rapid Miner tools, which set the role to determine the role of the attributes in the dataset. In addition, there is cross-validation to evaluate [26]. The Cross-Validation process can be seen in Figure 6.

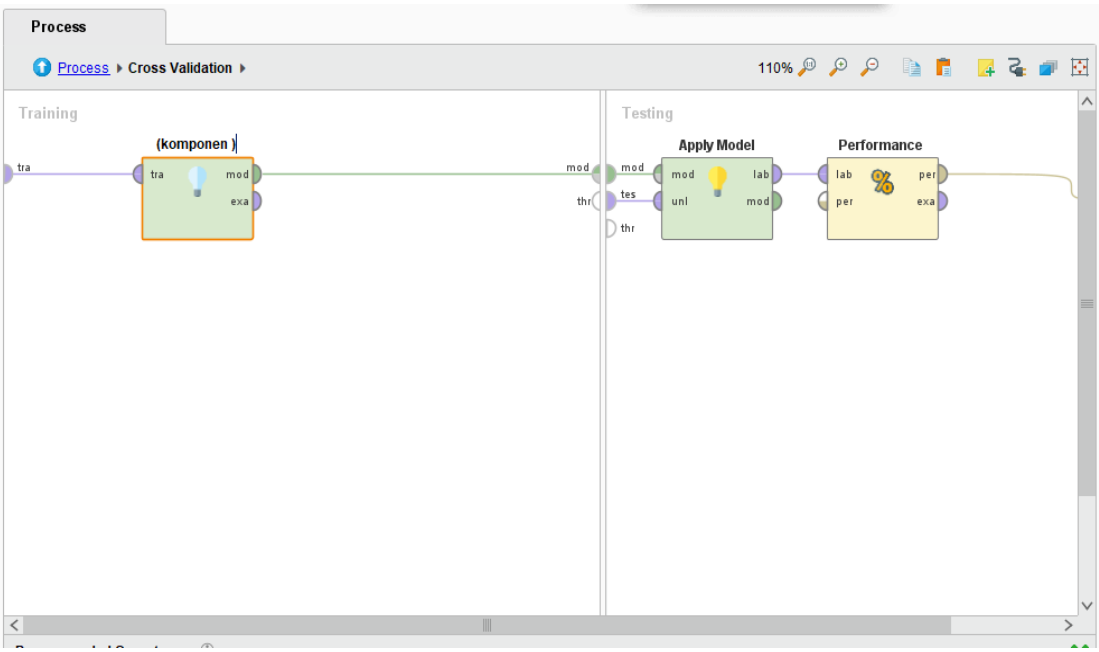


Figure 6. *Cross-Validation* Process

Figure 6 is a step for cross-validation which is an action that researchers take to find the accuracy of each method by dividing training and testing data in the form of applied model and performance [27].

3.2.Method Results

3.2.1. Naive Bayes Algorithm Results

The results of the Naïve Bayes method that has been carried out. The results of the study using Naive Bayes can be seen in Figure 7.

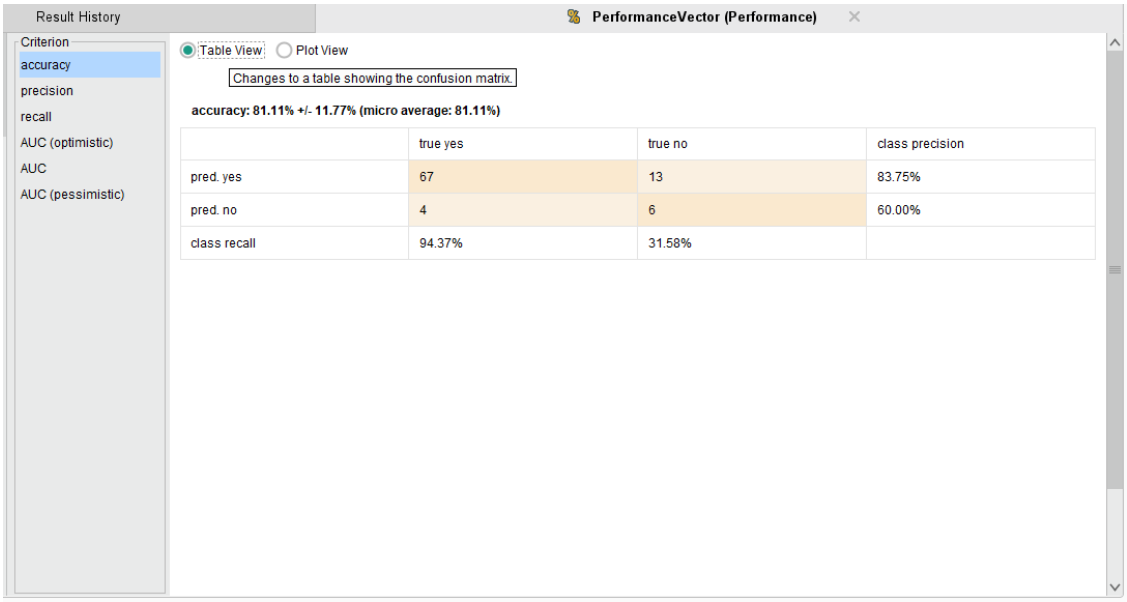


Figure 7. Accuracy Naive Bayes results

The results of the Naive Bayes method state that the accuracy rate of this method is 81.11%. Where is class precision for pred. yes is 83.75%, pred no is 60.00%. Besides that, there is also a performance vector for the Naïve Bayes method. However, there is also a plot view of this algorithm that can be viewed in Figure 8 dan data classification result in Figure 9.

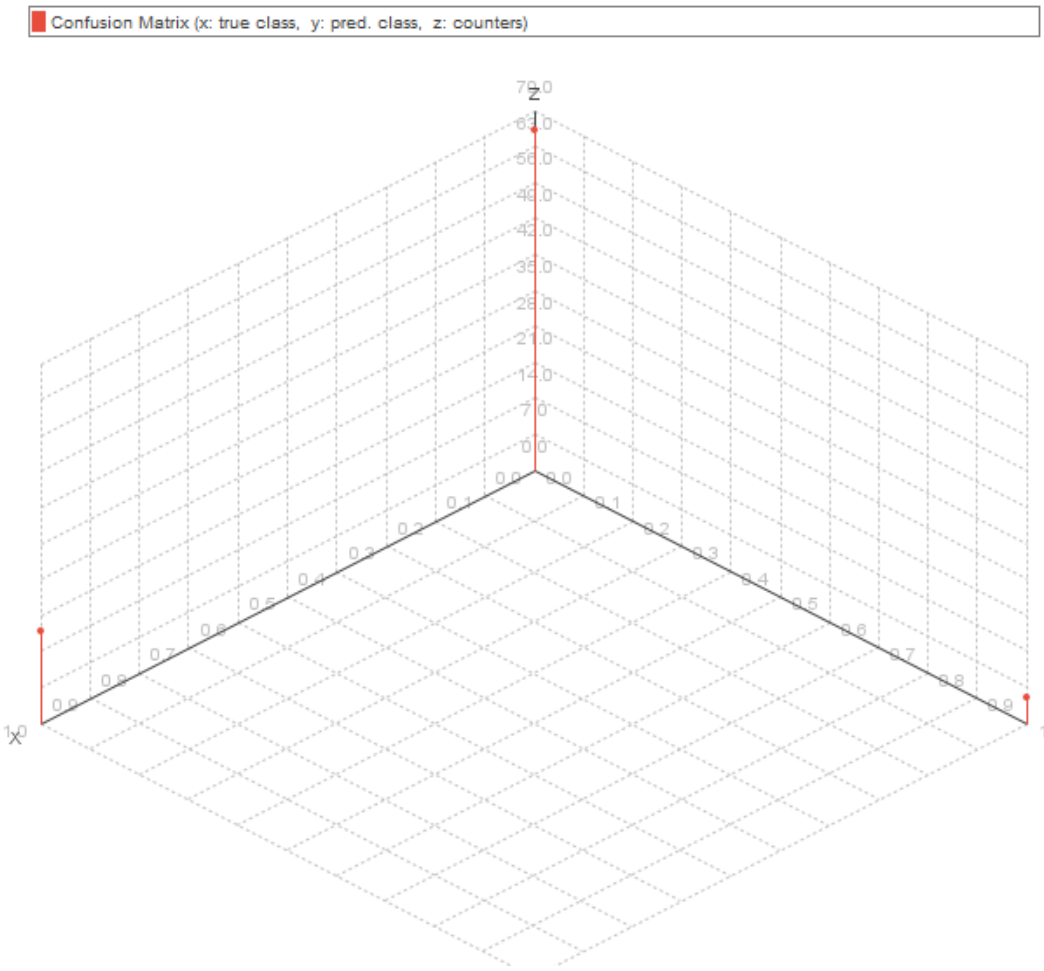


Figure 8. Plot View of the Naive Bayes Method

PerformanceVector

```
PerformanceVector:
accuracy: 81.11% +/- 11.77% (micro average: 81.11%)
ConfusionMatrix:
True:   yes     no
yes:    67      13
no:     4       6
precision: 60.00% (positive class: no)
ConfusionMatrix:
True:   yes     no
yes:    67      13
no:     4       6
recall: 30.00% +/- 25.82% (micro average: 31.58%) (positive class: no)
ConfusionMatrix:
True:   yes     no
yes:    67      13
no:     4       6
AUC (optimistic): 0.634 +/- 0.336 (micro average: 0.634) (positive class: no)
AUC: 0.634 +/- 0.336 (micro average: 0.634) (positive class: no)
AUC (pessimistic): 0.634 +/- 0.336 (micro average: 0.634) (positive class: no)
```

Figure 9. The data classification result using Naive Bayes method.

3.2.2. Decision Tree

The results of the Decision Tree method that has been carried out and can be viewed in Figure 10.

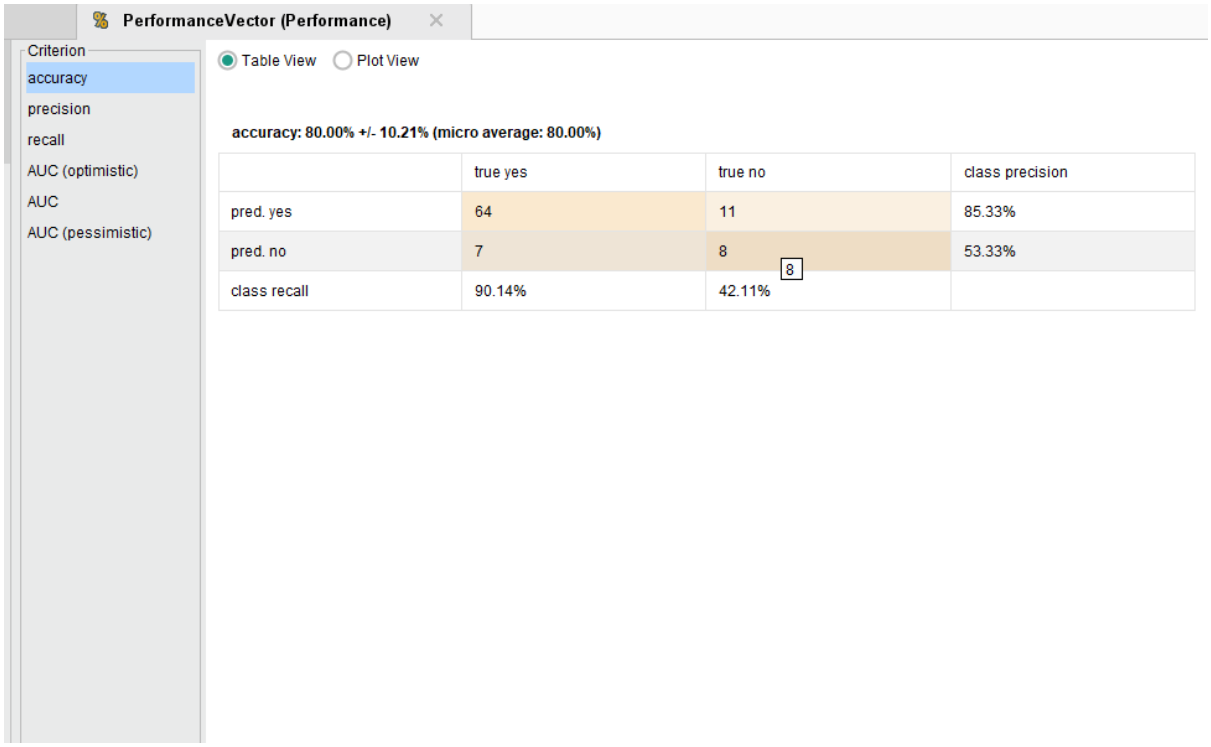


Figure 10. The result of Decision Tree accuracy.

The results of the Decision Tree method state that the accuracy rate of this method is 80.00%. Where is class precision for pred. yes is 85.33%, pred no is 53.33%. Besides that, there is also a performance vector for the Naïve Bayes method. However, there is also a plot view of this algorithm that can be viewed in Figure 11 and data classification result in Figure 12.

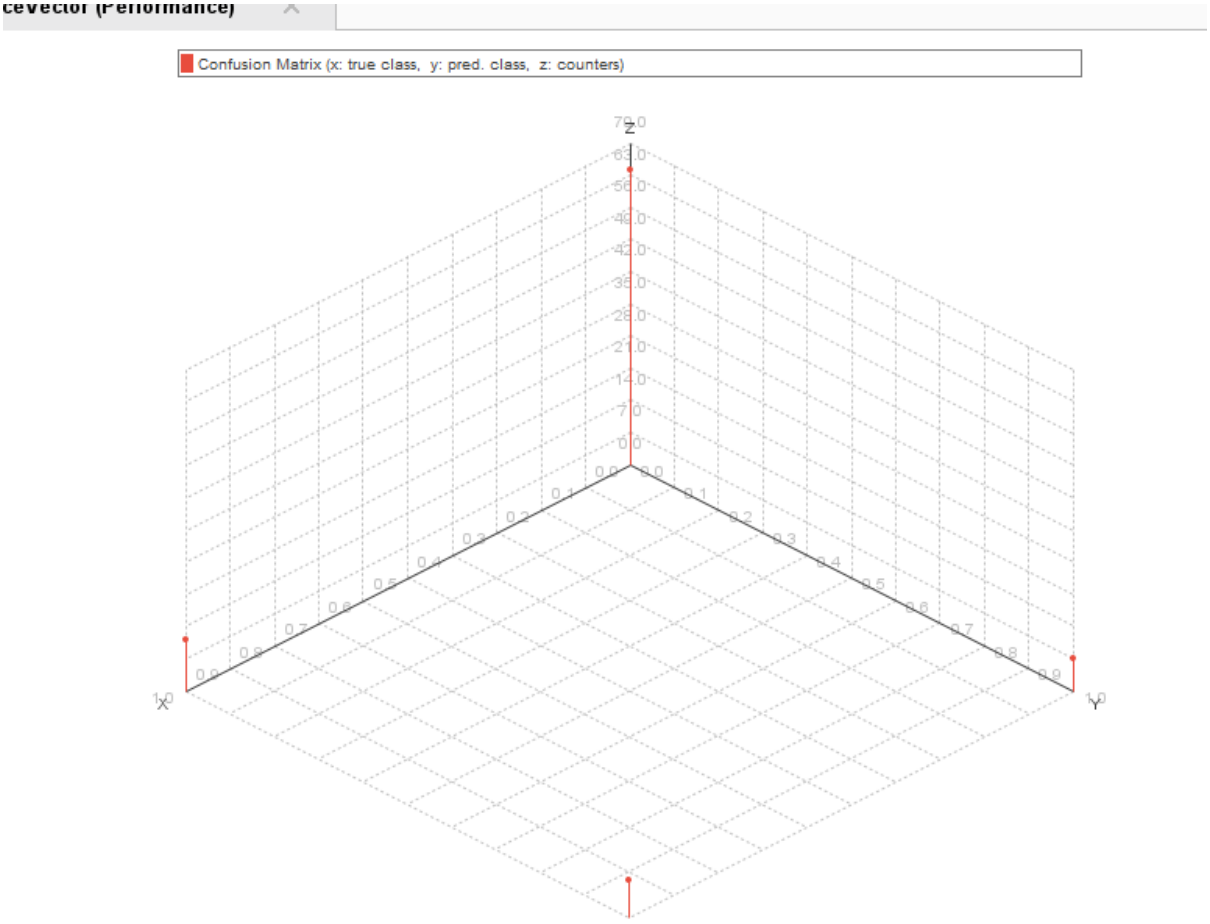


Figure 11. Plot View of the Decision Tree method.

PerformanceVector

```
PerformanceVector:
accuracy: 80.00% +/- 10.21% (micro average: 80.00%)
ConfusionMatrix:
True:   yes     no
yes:    64      11
no:     7       8
precision: 53.33% (positive class: no)
ConfusionMatrix:
True:   yes     no
yes:    64      11
no:     7       8
recall: 40.00% +/- 31.62% (micro average: 42.11%) (positive class: no)
ConfusionMatrix:
True:   yes     no
yes:    64      11
no:     7       8
AUC (optimistic): 0.871 +/- 0.120 (micro average: 0.871) (positive class: no)
AUC: 0.700 +/- 0.231 (micro average: 0.700) (positive class: no)
AUC (pessimistic): 0.564 +/- 0.371 (micro average: 0.564) (positive class: no)
```

Figure 12. The data classification result using Decision Tree method.

Performance vector itself is a form of description of the analysis results table. It's just that there are additions such as AUC to the performance vector of this Decision Tree.

3.2.3. K-NN

The results of the K-Nearest Neighbor method that has been carried out in Figure 13.

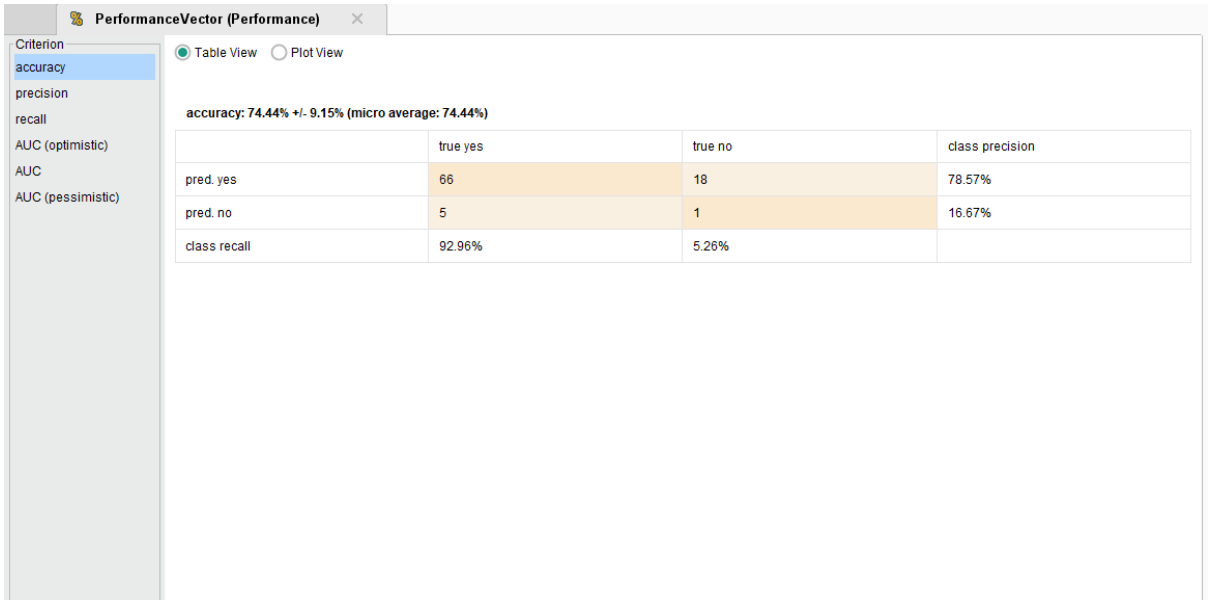


Figure 13. The result of K-Nearest Neighbor accuracy

The results of the K-Nearest Neighbor method state that the accuracy rate of this method is 74.44%. Where is class precision for pred. yes is 78.57%, while for pred no is 16.67%. Besides that, there is also a performance vector for the Naïve Bayes method. However, there is also a plot view of this algorithm that can be viewed in Figure 14 and data classification result in Figure 15.

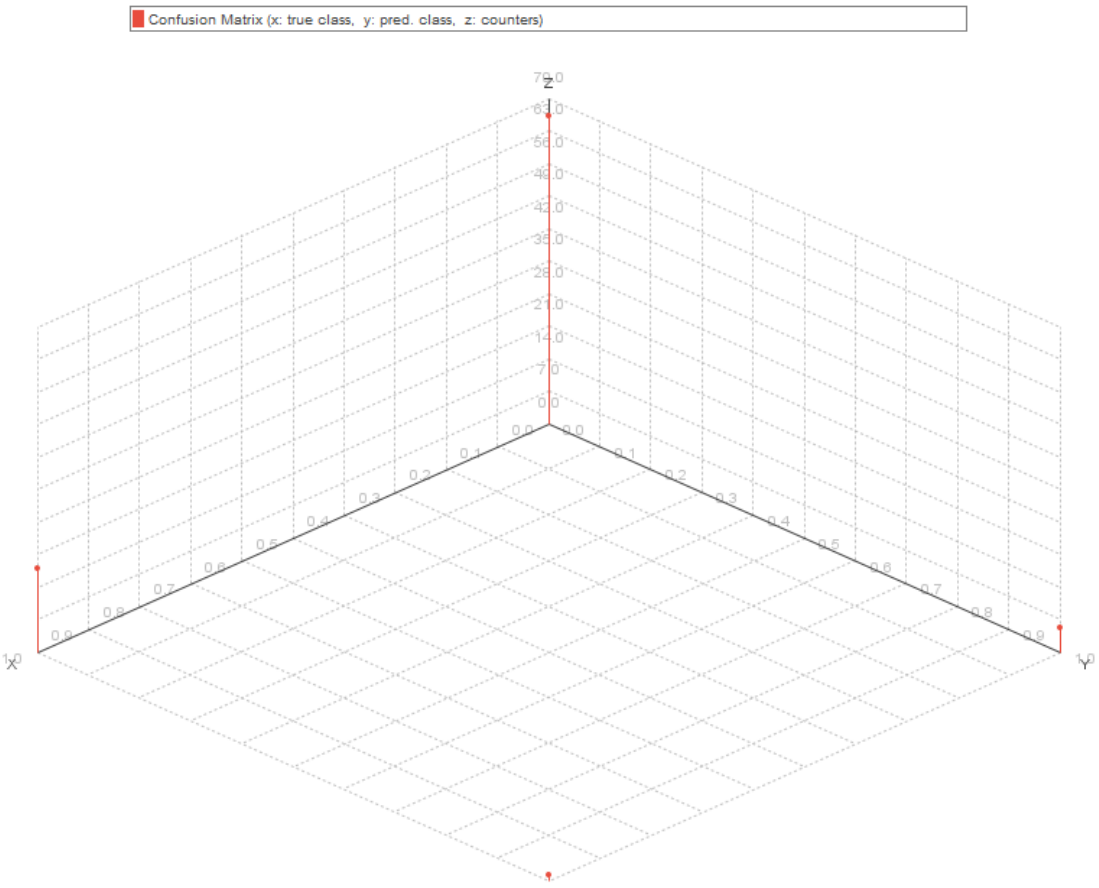


Figure 14. Plot View of K-Nearest Neighbor method

PerformanceVector

```
PerformanceVector:
accuracy: 74.44% +/- 9.15% (micro average: 74.44%)
ConfusionMatrix:
True:    yes     no
yes:     66      18
no:      5       1
precision: 16.67% (positive class: no)
ConfusionMatrix:
True:    yes     no
yes:     66      18
no:      5       1
recall: 5.00% +/- 15.81% (micro average: 5.26%) (positive class: no)
ConfusionMatrix:
True:    yes     no
yes:     66      18
no:      5       1
AUC (optimistic): 0.684 +/- 0.248 (micro average: 0.684) (positive class: no)
AUC: 0.641 +/- 0.278 (micro average: 0.641) (positive class: no)
AUC (pessimistic): 0.598 +/- 0.318 (micro average: 0.598) (positive class: no)
```

Figure 15. The data classification result using K-Nearest Neighbor method

3.2.4. Comparison of The Result

After testing the classification using the K-Nearest Neighbor, Naive Bayes, and Decision Tree methods was completed, the results obtained for the accuracy of the comparison of the three methods in carrying out the classification using the immunotherapy dataset. It was known in the previous section that the results obtained from testing the three methods received different results and were not too significant. [16] Table 1 displays a comparison of the outcomes as shown below.

Table 1. Comparison of the result

Method	Prediction Accuracy (%)
Naive Bayes	81.11
Decision Tree	80.00
K-Neares Neighbor	74.44

From the accuracy comparison above, it is known that the Naive Bayes algorithm is the most effective algorithm with the highest accuracy value of 81.11%.

4. Conclusion

Based on testing on the three methods used in this study, different results were obtained with differences in accuracy results that were not far from the processing of Immunotherapy Dataset data. Based on the test results of the naïve bayes method, decision tree, and K-Nearest Neighbor obtained accuracy rates of 81.11%, 80.00%, and 74.44%. From the results of tests conducted on the handling of data from the Immunotherapy Dataset, it shows that the naïve bayes method obtained more recommended results when compared to the Decision Tree method, and K-Nearest Neighbor.

REFERENCES

[1] A. Rahma, D. Claudia, F. A. Yulianto, and N. Romadhona, "Systematical Review : Pengaruh Olahraga Sepeda terhadap Penurunan Berat Badan Pada Dewasa Muda," *J. Integr. Kesehat. Sains*, vol. 3, no. 1, pp. 117–123, Mar. 2021, doi: 10.29313/jiks.v3i1.7427.

[2] C. J. L. Murray *et al.*, "Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019," *Lancet*, vol. 396, no. 10258, pp. 1223–1249, Oct. 2020, doi: 10.1016/S0140-6736(20)30752-2.

[3] F. Khozeimeh *et al.*, "Intralesional immunotherapy compared to cryotherapy in the treatment of warts," *Int. J. Dermatol.*, vol. 56, no. 4, pp. 474–478, Apr. 2017, doi: 10.1111/ijd.13535.

[4] N. C. Clarissa and I. Defiana, "Rumah Singgah untuk Penderita Kanker dengan Konsep Healing Environment," *J. Sains dan Seni ITS*, vol. 7, no. 2, Feb. 2019, doi: 10.12962/j23373520.v7i2.35295.

[5] A. Dwiantoro, I. Maulana, N. P. Damayanti, and R. N. Al Zahra, "Artificial intelligence (AI)

imaging for enhancement of parking security," *J. Student Res. Explor.*, vol. 1, no. 1, pp. 15–20, Dec. 2022, doi: 10.52465/josre.v1i1.110.
[6] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, A. Khosravi, P. Layegh, and S. Nahavandi, "An expert system for selecting wart treatment method," *Comput. Biol. Med.*, vol. 81, pp. 167–175, Feb. 2017, doi: 10.1016/j.compbiomed.2017.01.001.
[7] A. Supriyatna and W. P. Mustika, "Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, p. 152, Sep. 2018, doi: 10.30645/j-sakti.v2i2.78.
[8] G. A. W. SARI, "Peningkatan Akurasi Algoritma C4.5 Menggunakan Diskritisasi Dan Pemilihan Fitur Berbasis Korelasi," Universitas Gadjah Mada, 2020.
[9] Y. A. Pratama, "Analisis Metode Seleksi Fitur untuk Meningkatkan Akurasi pada Variant Metode Klasifikasi K-Nearest Neighbor (kNN)," Universitas Sumatera Utara, 2019.
[10] A. R. Safitri and M. A. Muslim, "Improved Accuracy of Naive Bayes Classifier for Determination of Customer Churn Uses SMOTE and Genetic Algorithms," *J. Soft Comput. Explor.*, vol. 1, no. 1, Sep. 2020, doi: 10.52465/joscex.v1i1.5.
[11] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote dan K-Nearest Neighbor," *J. Inf. Syst. Dev.*, vol. 3, no. 1, pp. 44–49, 2018.
[12] M. H. Ramdani, I. G. P. S. Wijaya, and R. Dwiyansaputra, "Optimalisasi Pengenalan Wajah Berbasis Linear Discriminant Analysis Dan K-Nearest Neighbor Menggunakan Particle Swarm Optimization," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 4, no. 1, pp. 40–51, 2022.
[13] L. I. Prahartiwi and W. Dari, "Komparasi Algoritma Naive Bayes, Decision Tree dan Support Vector Machine untuk Prediksi Penyakit Kanker Payudara," *J. Tek. Komput.*, vol. 7, no. 1, pp. 51–54, Feb. 2021, doi: 10.31294/jtk.v7i1.9191.
[14] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, Dec. 2021, doi: 10.32493/informatika.v5i4.7622.
[15] R. Y. Hayuningtyas, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 6, no. 1, pp. 18–22, Apr. 2019, doi: 10.31294/ji.v6i1.4685.
[16] U. Erdiansyah, A. Irmansyah Lubis, and K. Erwansyah, "Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, p. 208, Jan. 2022, doi: 10.30865/mib.v6i1.3373.
[17] A. Nurdina and A. B. I. Puspita, "Naive Bayes and KNN for Airline Passenger Satisfaction Classification: Comparative Analysis," *J. Inf. Syst. Explor. Res.*, vol. 1, no. 2, pp. 83–92, 2023, doi: https://doi.org/10.52465/joiser.v1i2.167.
[18] R. Harun, K. C. Pelangi, and Y. Lasena, "Penerapan Data Mining untuk Menentukan Potensi Hujan Harian dengan Menggunakan Algoritma K Nearest Neighbor (KNN)," *J. Manaj. Inform. dan Sist. Inf.*, vol. 3, no. 1, pp. 8–15, 2020.
[19] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *J. Ilm. Matrik*, vol. 22, no. 1, pp. 73–82, Mar. 2020, doi: 10.33557/jurnalmatrik.v22i1.843.
[20] M. R. Handoko and N. Neneng, "Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 50–58, 2021.
[21] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, "Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 331–337, Dec. 2021, doi: 10.47065/bits.v3i3.1093.
[22] L. Alfaris, R. C. Siagian, A. C. Muhammad, U. I. Nyuswantoro, N. Laeiq, and F. D. Mobo, "Classification of Spiral and Non-Spiral Galaxies using Decision Tree Analysis and Random Forest Model: A Study on the Zoo Galaxy Dataset," *Sci. J. Informatics*, vol. 10, no. 2, pp. 139–150, May 2023, doi: 10.15294/sji.v10i2.44027.
[23] V. Anestiviya and A. F. O. Pasaribu, "Analisis Pola Menggunakan Metode C4.5 untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : SMAN 1 Natar)," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 80–85, 2021.

[24] S. Febriani and H. Sulistiani, "Analisis Data Hasil Diagnoga Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
[25] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, and P. Layegh, "Immunotherapy Dataset," *UCI Mach. Learn. Repos.*, 2018, doi: https://doi.org/10.24432/C5DC72.
[26] A. M. Siregar and A. Puspabhuana, *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*, Pertama. Surakarta: CV Kekata Group, 2017.
[27] A. Y. Rangan, M. I. Ukkas, and S. Qomariah, "Implementasi Metode Support Vector Mechine dan K-Means Clustering dengan Support Vector Mechine Tema Karya Ilmiah pada Stmik Widya Cipta Dharma," vol. 3, pp. 274–279, 2018.